Security For Open Science –
Proposed Addition to the Federal Plan to Secure CyberSpace

Deb Agarwal (DAAgarwal@lbl.gov)
Brian Tierney (BLTierney@lbl.gov)

Lawrence Berkeley National Laboratory
1 Cyclotron Rd., MS 50B-2239, Berkeley, CA 94720

Deborah Frincke (deborah.frincke@pnl.gov)

Pacific Northwest National Laboratory
902 Battelle Blvd, MSIN: K7-30
Richland, WA 99352

Abstract

The United States has developed and operates some of the world's most advanced research and development, open scientific research facilities. Every year these facilities are used by more than 18,000 researchers from universities, government agencies, and private industry worldwide, to access technologies and instrumentation available nowhere else. Many of these researchers rarely, if ever, visit the facility they are using.

Unfortunately, the openness required for effective scientific collaboration leads to increased vulnerability. A recent large-scale attack on several national supercomputing centers provides an example of the importance of this challenge. The attack took the San Diego Supercomputer Center off the network for an entire week, and other national facilities were likewise disrupted for several weeks. There are many challenges left to be addressed before we have an easy-to-use, comprehensive defense-in-depth protection strategy for open science at all levels – and the importance of the facilities, and unique nature of the challenges of securing open science deserves explicit consideration in the Federal Plan. We therefore suggest that the national plan be expanded to include a focus on this element, and that a national research agenda be devised and funded, with the intent of identifying and integrating the needs and efforts of stakeholders such as the Department of Energy, National Science Foundation, Department of Homeland Security, and universities worldwide. It is in the national interest to continue the existing tradition of maintaining a strong, and *open*, suite of resources so that US scientists and others can intermingle; however, it would be naïve to assume that existing strategies to maintain such facilities will be both effective and usable without focused and integrated advances in research on open science security questions, and transitioning of this research into the facilities and scientific communities themselves.

Security Issues for Open Science

The Department of Energy (DOE) Office of Science, the National Science Foundation (NSF), and the National Institutes of Heath (NIH) are responsible for the operation of some of the nation's most advanced research and development user facilities located at the national laboratories and universities. These state-of-the-art facilities are shared with the science community worldwide, and contain technologies and instrumentation that are available nowhere else. For example, the National Synchrotron Light Source at Brookhaven National Laboratory is the world's brightest continuous source of X-rays and ultraviolet radiation for research. The Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory houses one of the world's most powerful widebore nuclear magnetic resonance (NMR) spectrometers. The Spallation Neutron Source, at Oak Ridge National Laboratory, will provide the most intense pulsed neutron beams in the world for scientific research and industrial development. Other DOE user facilities include the Advanced Photon Source, the National Energy Research Scientific Computing Center (NERSC), and the new National Leadership Computing Facility. NSF facilities include the San Diego Supercomputer Center, the National Center for Supercomputing Applications, the National Center for Atmospheric Research, and the National Laboratory for Applied Network Research. Each year, the DOE Office of Science facilities are used by more than 18,000 researchers from universities, other government agencies, and private industry. NSF supercomputer centers also provide compute resources to researchers around the world.

Although many of the major science experiments in which DOE and NSF researchers participate are located at DOE and NSF facilities, increasingly, key experiments are located elsewhere, with facilities spanning international boundaries. The CMS and ATLAS experiments at the Large Hadron Collider (LHC) to be built at CERN in Switzerland each involve approximately 2000 physicists from around the world. DOE, as the host of the U.S. CMS and ATLAS Tier 1 centers, is providing a key element of the global support infrastructure for these experiments. NSF is providing key infrastructure to these experiments by sponsoring universities to host the Tier 2 centers. The ITER fusion reactor will be located in France and will involve fusion scientists from all over the world in its operations and experiments. These DOE and NSF user facilities and research collaborations are valued at billions of dollars and have extensive computing and networking resources. Protecting the infrastructure and the experiments tied up in this investment will require collaboration and cooperation between cybersecurity personnel in the US, Switzerland, France and other collaborating nations. Additional research is needed to support such broad scale interactions.

The high performance environment, global user population, and diversity of custom applications and software in widespread open science environments relied upon by efforts such as ATLAS make protecting the facilities and detecting malicious attacks challenging. Enabling first-class participation in this kind of environment requires continual vigilance. As seen in the previously mentioned attacks on the US' supercomputing facilities, recovery of a high-value user facility can take many days or weeks, during which the facility is unavailable for its mission. DOE and NSF facilities and research collaborations can ill afford to be offline for extended periods due to security incidents.

Research investments in defense or commercial systems will not be sufficient to meet the needs of open science. For example, in an open science environment, traditional cybersecurity mechanisms aimed at protecting perimeters and keeping information locked within a restricted area, such as a border firewalls, are of limited value. Instead, the open science environment demands a defense in depth approach with a default allow policy; only recognized malicious activity is automatically blocked at the border, and other protections are used to defend against novel attacks. Otherwise, new forms of scientific collaboration will be slowed or prevented, as they cross administrative boundaries. This environment therefore requires much stronger internal protections. It will be important also that these pervasive protections, intrusion detection, and data analysis capabilities be able to operate in a global environment that includes petascale computing and 10-40 Gbit networks, and unique SCADA facilities. Development and deployment of the cybersecurity mechanisms specifically geared towards protecting science resources and participation in distributed science projects is needed. Addressing the cybersecurity needs of open science has already led to breakthroughs in cybersecurity research and development such as the *Bro* intrusion detection system at LBNL and the 24x7 security system at PNNL.

Another example of where open science differs from traditional commercial or government operations can be seen in the distribution of software. Software applications in open science environments often open ports, transfer files, and coordinate activities across sites. One example of this largely non-commercial software is the Grid software, which incorporates authentication, authorization, scheduling, data transfer, portals, etc. Grid software forms the core of the Open Science Grid, which is used by many science collaborations (often referred to as *virtual organizations*) including the Atlas and CMS experiments at the LHC. Within these virtual organizations, the authentication and authorization are federated to enable cross-site authentication, incorporate dynamically available resources, manage allocation of resources, and track individuals as they access the sites and resources of the virtual organization. This virtual organization model redefines the traditional enclave into one that crosses and incorporates many individual site borders and includes personnel and resources from a wide range of sites. Within the confines of the virtual organization, large quantities of data need to be able to move using high-speed communication links from site to site. Tools and services to allow virtual organizations to better monitor their resources and perform incident containment are needed.

The Department of Energy national laboratories and large NSF centers are in need of better cybersecurity mechanisms built for the open science environment. In addition, these facilities provide an ideal environment in which to build, test, and deploy a cooperative cybersecurity system. There is a unique level of consensus and information interchange across and within enclaves. This is due to the fact that the labs and their wide-area network are all under the Department of Energy. A similar situation exists within the NSF centers. It is also aided by the fact that many of the science programs are based on large collaborations that span many enclaves and cross national borders. Projects such as the Open Science Grid span both DOE and NSF facilities and networks. These provide an ideal environment for deploying and testing interoperable cybersecurity systems across enclaves with different levels of trust.

A white paper describing the cybersecurity needs of open science in greater detail is available at http://www.dsd.lbl.gov/~deba/publications/DOE-NSF-Cybersecurity-white-paper-draft-v3.1.pdf